

ACM SIGCOMM Workshop on Big Data Analytics and Machine Learning for Data Communication Networks

*Original*

ACM SIGCOMM Workshop on Big Data Analytics and Machine Learning for Data Communication Networks / Casas, P.; Mellia, M.; Dainotti, A.; Zseby, T.. - ELETTRONICO. - 1:(2017), pp. 1-1. (Intervento presentato al convegno Big Data Analytics and Machine Learning for Data Communication Networks tenutosi a Los Angeles nel August 2017).

*Availability:*

This version is available at: 11583/2751335 since: 2019-09-11T12:07:10Z

*Publisher:*

Association for Computing Machinery, Inc

*Published*

DOI:

*Terms of use:*

openAccess

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

ACM postprint/Author's Accepted Manuscript

(Article begins on next page)

August 21, 2017  
Los Angeles, CA, USA



Association for  
Computing Machinery

*Advancing Computing as a Science & Profession*



# Big-DAMA '17

Proceedings of the 2017

Workshop on Big Data Analytics and Machine  
Learning for Data Communication Networks

*Sponsored by:*

**ACM SIGCOMM**

*Part of:*

**SIGCOMM '17**



**Association for  
Computing Machinery**

*Advancing Computing as a Science & Profession*

**The Association for Computing Machinery**  
2 Penn Plaza, Suite 701  
New York, New York 10121-0701

Copyright © 2017 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from [permissions@acm.org](mailto:permissions@acm.org) or Fax +1 212 869-0481.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through [www.copyright.com](http://www.copyright.com).

**Notice to Past Authors of ACM-Published Articles**

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that has been previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform [permissions@acm.org](mailto:permissions@acm.org), stating the title of the work, the author(s), and where and when published.

**ISBN: 978-1-4503-5054-9**

Additional copies may be ordered prepaid from:

**ACM Order Department**  
PO Box 30777  
New York, NY 10087-0777, USA

Phone: +1 800 342-6626 (USA and Canada)

+1 212 626-0500 (Global)

Fax: +1 212 944-1318

Email: [acmhelp@acm.org](mailto:acmhelp@acm.org)

Hours of Operation: 8:30 am–4:30 pm ET

# ACM SIGCOMM Workshop on Big Data Analytics and Machine Learning for Data Communication Networks

Big-DAMA 2017

## CHAIR'S WELCOME

The explosion in volume and heterogeneity of data communication network measurements opens the door to the massive application of machine learning and artificial intelligence technology in networking. While machine learning is today systematically and successfully applied in many other data-driven domains, its application is in an infancy stage of development in the networking domain. The ACM SIGCOMM Workshop on Big Data Analytics and Machine Learning for Data Communication Networks, **Big-DAMA**, fosters the research and development of novel analytical approaches and technical solutions that can exploit Big Data technology in the analysis of complex communication networks such as the Internet.

**Big-DAMA 2017 accepted 9 technical papers out of 25**, high-quality submissions. The paper review process included an evaluation phase by PC members, followed by an online discussion and a subsequent shepherding phase on selected papers. The resulting program features a variety of high-quality papers focusing on different aspects of machine learning and network measurements, including network security, network performance, big-data monitoring and user characterization. Big-DAMA 2017 also features three exciting talks from recognized researchers and practitioners in machine learning and network measurements domains, including:

### What I Learned about Big Data Working with Biologists, Neuroscientists and Climate Scientists

Constantine Dovrolis (Georgia Tech)

### Scaling BGP Big Data for Network Operations, SDN and Research

Tim Evens (Cisco)

### Big Data Begets Big Data: Understanding Modern Datacenter Networks

Alex C. Snoeren (UC San Diego)

We believe that the application of machine learning and Big Data technology in the analysis of data communication networks has a major future, and we hope that all attendees will enjoy the excellent program and find a nice and constructive environment to discuss on this bright future.

**Pedro Casas**

Austrian Institute of Technology

**Marco Mellia**

Politecnico di Torino

**Alberto Dainotti**

CAIDA, UC San Diego

**Tanja Zseby**

TU Wien

## WORKSHOP ORGANIZATION

### Workshop Chairs

Pedro Casas, *Austrian Institute of Technology, Austria*

Marco Mellia, *Politecnico di Torino, Italy*

Alberto Dainotti, *CAIDA UC San Diego, USA*

Tanja Zseby, *TU Wien, Austria*

### Program Committee

Isabel Amigo, *Télécom Bretagne, France*

Arian Bär, *BMW Group, Germany*

Elena Baralis, *Politecnico di Torino, Italy*

Pere Barlet-Ros, *UPC BarcelonaTech, Spain*

Alessandro D'Alconzo, *Austrian Institute of Technology, Austria*

Constantine Dovrolis, *Georgia Tech, USA*

Idilio Drago, *Politecnico di Torino, Italy*

Pierdomenico Fiadino, *EURECAT, Spain*

Marco Fiore, *CNR-IEIT, Italy*

Tobias Flach, *Google, USA*

Kensuke Fukuda, *NII Tokyo, Japan*

Félix Iglesias Vázquez, *TU Wien, Austria*

Hyunchul Kim, *Seoul National University, South Korea*

Federico Larroca, *Universidad de la República, Uruguay*

Matthew Luckie, *University of Waikato, New Zealand*

Ian Marsh, *RISE SICS, Sweden*

Pietro Michiardi, *EURECOM, France*

Philippe Owezarski, *CNRS, France*

Fabio Ricciato, *University of Ljubljana, Slovenia*

Dario Rossi, *Télécom ParisTech, France*

Narseo Vallina Rodriguez, *IMDEA Networks, Spain*

# Contents

<b>Ensemble-learning Approaches for Network Security and Anomaly Detection</b> . . . . .	1
Juan Vanerio ( <i>Universidad de la República/AIT Austrian Institute of Technology</i> ); Pedro Casas ( <i>AIT Austrian Institute of Technology</i> )	
<b>Cluster-Based Load Balancing for Better Network Security</b> . . . . .	7
Gal Frishman ( <i>IDC</i> ); Yaniv Ben-Itzhak, Oded Margalit ( <i>IBM Research</i> )	
<b>Net2Vec: Deep Learning for the Network</b> . . . . .	13
Roberto Gonzalez, Filipe Manco, Alberto Garcia-Duran, Jose Mendes, Felipe Huici, Saverio Niccolini, Mathias Niepert ( <i>NEC Labs. Europe</i> )	
<b>o'zapft is: Tap Your Network Algorithm's Big Data!</b> . . . . .	19
Andreas Blenk, Patrick Kalmbach ( <i>Technical University of Munich</i> ); Stefan Schmid ( <i>Aalborg University</i> ); Wolfgang Kellerer ( <i>Technical University of Munich</i> )	
<b>Hierarchical IP flow clustering</b> . . . . .	25
Kamal Shadi ( <i>Georgia Institute of Technology</i> ); Preethi Natarajan ( <i>Cisco</i> ); Constantine Dovrolis ( <i>Georgia Institute of Technology</i> )	
<b>NETPerfTrace – Predicting Internet Path Dynamics and Performance with Machine Learning</b> . . .	31
Sarah Wassermann ( <i>Université de Liège</i> ); Pedro Casas ( <i>AIT Austrian Institute of Technology</i> ); Thibaut Cuvelier, Benoit Donnet ( <i>Université de Liège</i> )	
<b>Neural Network Based Wavelength Assignment in Optical Switching</b> . . . . .	37
Craig Gutterman ( <i>Columbia University</i> ); Weiyang Mo, Shengxiang Zhu, Yao Li, Daniel C. Kilper ( <i>University of Arizona</i> ); Gil Zussman ( <i>Columbia University</i> )	
<b>Call Detail Records for Human Mobility Studies: Taking Stock of the Situation in the "Always Connected Era"</b> . . . . .	43
Pierdomenico Fiadino, Víctor Ponce-López ( <i>EURECAT</i> ); Juan Antonio Torrero-Gonzalez ( <i>ORANGE SPAIN</i> ); Marc Torrent-Moreno ( <i>EURECAT</i> ); Alessandro D'Alconzo ( <i>AIT</i> )	
<b>Users' Fingerprinting Techniques from TCP Traffic</b> . . . . .	49
Luca Vassio, Danilo Giordano, Martino Trevisan, Marco Mellia ( <i>Politecnico di Torino</i> ); Ana Paula Couto da Silva ( <i>Universidade Federal de Minas Gerais</i> )	